D-CoRP: Differentiable Connectivity Refinement for Functional Brain Networks

Haoyu Hu[1], Hongrun Zhang[1,2], and Chao Li[1,3*]

[1] Department of Clinical Neurosciences, University of Cambridge
[2] CRUK Cambridge Institute, University of Cambridge
[3] Department of Applied Maths and Theoretical Physics, University of Cambridge
{cl647@cam.ac.uk}

Abstract. Brain network is an important tool for understanding the brain, offering insights for scientific research and clinical diagnosis. Existing models for brain networks typically primarily focus on brain regions or overlook the complexity of brain connectivities. MRI-derived brain network data is commonly susceptible to connectivity noise, underscoring the necessity of incorporating connectivities into the modeling of brain networks. To address this gap, we introduce a differentiable module for refining brain connectivity. We develop the multivariate optimization based on information bottleneck theory to address the complexity of the brain network and filter noisy or redundant connections. Also, our method functions as a flexible plugin that is adaptable to most graph neural networks. Our extensive experimental results show that the proposed method can significantly improve the performance of various baseline models and outperform other state-of-the-art methods, indicating the effectiveness and generalizability of the proposed method in refining brain network connectivity. The code is available at https://github.com/Fighting-HHY/D-CoRP

Keywords: Functional Brain Network · Graph Structure Learning · Information Bottleneck.

1 Introduction

The human brain comprises multiple sophisticated brain regions, empowering it to navigate intricate cognitive processes and handle everyday tasks [13]. Recent advances in neuroscience suggest that brain function is fulfilled by the coordinated activities of brain regions through connectivities [17]. Brain networks, graphical representations of the brain, hold the potential to characterize regional and global brain activities [5,14,27]. Functional brain networks, derived from function MRI (fMRI), is one of the most important tools for characterizing the brain function. However, it remains challenging to effectively leverage the information in functional brain networks due to their non-Euclidean structure.

---

* corresponding author

Graph neural networks (GNNs) [33] promise to offer effective toolkits to analyze graph data [25]. For example, BrainGNN is proposed to effectively aggregate the information from functional brain networks for learning tasks [15,26]. Graph pooling methods are developed for characterizing network modularity, e.g., TopK, a score-based pooling method [9] and DiffPool, a hierarchical pooling method [30]. However, these models only focus on node features without considering edge connections, which may particularly limit the model performance on MRI-derived brain networks that are prone to edge noises or redundancies[28].

Previous studies proposed graph structure learning (GSL) to denoise and optimize graph structure [34]. Despite successes, existing GSL models are developed for specific types of graphs (e.g. community networks, population graphs) using rigid optimization approaches (e.g., community property optimization), and they are often limited in generalizability to other graphs once trained. Scanty models have been proposed for brain networks so far. Though a recent model [32] proposes to reconstruct brain networks through time windows extracting features from fMRI, it cannot work directly on graphs (e.g., correlation matrices).

Information bottleneck (IB) theory [20] employs a probabilistic optimization between predictive performance and information compression by focusing on model-processed features of the input graph. Unlike typical GSL methods, IB is not restricted to a specific type of graph, which brings the advantages of generalizability while effectively removing noisy and irrelevant information. These advantages could equip IB as a competitive framework for denoising graphs.

To denoise the graph, previous IB-based methods perform subgraph filtering that updates the graph structure, e.g., subgraph information bottleneck (SIB) [31]. However, this approach treats the entire subgraph as processed features without performing differentiable refinement, which may lead to less desirable performance in removing edges. By contrast, other methods perform edge and node filtering, preserving the overall structure, e.g., VIB-GSL [19]. Nevertheless, these methods assume the optimized graph follows a standard Gaussian distribution, which may over-simplify the true distributions of edge connections that vary across different brain regions. Finally, all these methods are implemented as a fixed framework, lacking adaptability to existing GNN frameworks and hindering wider applicability. Thus, it remains a challenge to develop additional processing steps to refine edge connections integrated into existing frameworks.

We propose an edge-denoising method for functional brain networks based on IB theory, namely **D**ifferentiable **Co**nnection **R**efinement **P**lugin (**D-CoRP**), serving to detect and remove noisy or irrelevant connections from brain networks as an adaptable plugin. Our main contributions include:

1) We design a new differentiable sampling method that renders the plugin learnable, considering the probability of completely removing noisy connections.

2) We propose a new optimization strategy to model each connection independently with multivariate distribution targets in an IB-adopted framework.

3) As far as we know, this is the first plugin for connectivity refinement of brain networks, adding minimal computation to backbone training. We also provide an efficient mode for flexible adoption.

We evaluated our plugin on three datasets of functional brain networks. D-CoRP is shown to effectively remove noisy and irrelevant connections and enhance the performance of the GNN backbones, achieving a better and more stable performance compared with other state-of-the-art methods.

2   Methods

2.1   Overview

Define the target brain network to be refined as $\boldsymbol{G}_t$. Suppose $\boldsymbol{G}_t = (\boldsymbol{X}, \boldsymbol{A})$, where $\boldsymbol{X} \in \mathbb{R}^{N \times F}$ is brain region features such as brain activities at each region or geometric properties of the brain network and $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix consisting of connections between each brain regions (e.g., for functional brain network derived from fMRI data, the connection can be Pearson correlation between two brain regions' activities), with N being the amount of brain region, F being the dimension of region features. We aim to find a refined functional brain network adjacency matrix $\boldsymbol{A}_{re}$ based on the original $\boldsymbol{A}$, formulated as,

$$\boldsymbol{A}_{re} = \boldsymbol{m} \odot \boldsymbol{A}, \tag{1}$$

where $\boldsymbol{m} \in \mathbb{R}^{N \times N}$ is the learnable edge mask with binary elements, indicating the essential connections within brain networks. To learn the refinement mask $\boldsymbol{m}$, we propose to apply the IB theory [20,21], which makes $\boldsymbol{m}$ as a processed feature connected to the model's input and output.

2.2   Differentiable Connectivity Refinement

The binary refinement matrix $\boldsymbol{m}$ is naturally modeled as being Bernoulli distributed. However, the binary values are not differentiable, which hinders the gradient propagation in the learning process; thus, it is intractable. To address this, we relax the binary matrix $\boldsymbol{m}$ to a pseudo binary matrix, with each value in it being between $[0, 1]$, which can be uniformly sampled from a predefined function. We design this function such that a sampled value will be mostly closed to 0 or 1, resembling to the binary case (see Eq.(4) for the specific expression), which is parameterized by a learning probability matrix $\boldsymbol{\mathcal{P}}$, indicating the probabilities of corresponding edges being preserved.

Generation of Selection Probabilities (Distribution Estimator) The probability matrix $\boldsymbol{\mathcal{P}}$ is obtained via an auxiliary graph network (GAT) with the node features $\boldsymbol{X}$ and adjacency matrix $\boldsymbol{A}$, followed by the operation Sigmoid,

$$\boldsymbol{Z} = \text{GAT}(\boldsymbol{X}, \boldsymbol{A}). \tag{2}$$

$$\boldsymbol{\mathcal{P}} = \text{Sigmoid}\left(\boldsymbol{Z}\boldsymbol{Z}^T\right). \tag{3}$$

where $\boldsymbol{Z}, \boldsymbol{\mathcal{P}} \in \mathbb{R}^{N \times N}$. We choose GAT to distil critical graph information due to its efficient attention mechanism.
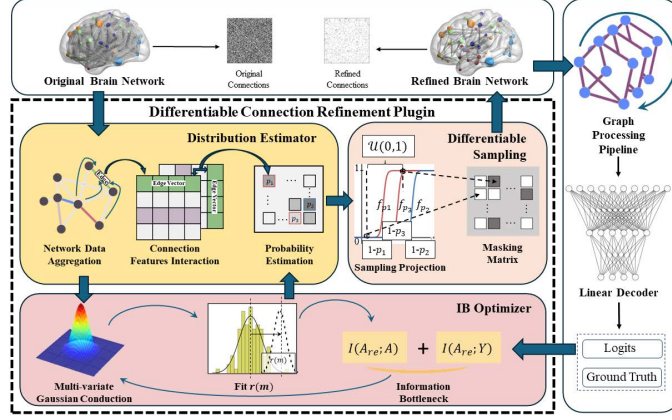
Fig. 1: **Model architecture.** D-CoRP refines the graph connection in two steps: 1) Differentiable sampling for the masking matrix (i.e.,**Distribution Estimator** extracts the necessity probability of network connections and **Differentiable Sampling** utilizes differentiable sampling methods filtrating necessary connections); 2) Information bottleneck optimization for the refined results as shown in **IB Optimizer** part.

Differentiable Sampling for Masking Matrix   The sampling is performed via a differentiable function adapted from the concrete relaxation of Bernoulli distribution [19]:

$$\mathcal{F}(\boldsymbol{\mathcal{P}}|\boldsymbol{A}) = \mathcal{F}(\boldsymbol{\mathcal{P}}, \boldsymbol{\Pi} : \mathcal{T}) = \frac{1}{2}(1 - \tanh(\frac{\boldsymbol{\Pi} + \boldsymbol{\mathcal{P}} - 1}{\mathcal{T}})) = \frac{1}{1 + e^{-\frac{2(\boldsymbol{\Pi} + \boldsymbol{\mathcal{P}} - 1)}{\mathcal{T}}}} \quad (4)$$

where $\boldsymbol{\Pi} \sim \mathcal{U}(0,1)$ provides uniformly random sampling. $\mathcal{T}$ represents the temperature for the function. This function can directly relate the probability matrix to the sampling process as Bernoulli distribution. When $\mathcal{T} \to 0$, we can treat $\mathcal{F}(\boldsymbol{\mathcal{P}}, \boldsymbol{\Pi} : \mathcal{T}) \approx \text{Ber}(\boldsymbol{\mathcal{P}})$, a Bernoulli distribution, while $\mathcal{F}(\boldsymbol{\mathcal{P}}, \boldsymbol{\Pi} : \mathcal{T})$ always keeps continuous in the $[0, 1]$ range. Unlike the previous concrete relaxation, Eq. (4) achieves the closed interval [0,1] for $\boldsymbol{\mathcal{P}}$, ensuring the full or zero probability of preserving brain connections. Also, it provides a smoother change of gradient without computing the logarithm, preventing the training process from exponential explosion.

2.3   Optimization with Information Bottleneck

In D-CoRP, the adjacency matrix $\boldsymbol{A}$ is the input data and $\boldsymbol{A}_{re}$ is the processed feature. Combined with general variational bounds developed by Alemi et al. [1], we design the optimization for brain network refinement (the specific derivation is provided in the supplementary):

$$\mathcal{L}_{IB} = \mathbb{E}_{\boldsymbol{A}_{re} \sim p_\theta(\boldsymbol{A}_{re}|\boldsymbol{A})}[-\log q(\boldsymbol{Y}|\boldsymbol{A}_{re})] + \beta \text{KL}[p(\boldsymbol{A}_{re}|\boldsymbol{A}), r(\boldsymbol{A}_{re})] \quad (5)$$

Further introducing the upper bound of Eq. (5)'s right side, we conduct two tractable optimization targets for brain network connectivity refinement:

$$\mathcal{L}_{AIB} = \mathbb{E}_{\boldsymbol{A}_{re} \sim p_\theta(\boldsymbol{A}_{re}|\boldsymbol{A})}[-\log q(\boldsymbol{Y}|\boldsymbol{A}_{re})] + \beta \mathrm{KL}[\mathcal{N}(\boldsymbol{\mathcal{P}};\boldsymbol{\Sigma}),\mathcal{N}(\boldsymbol{\mu_m};\boldsymbol{\Sigma})] \qquad (6)$$

$$\mathcal{L}_{AIB-E} = \mathbb{E}_{\boldsymbol{A}_{re} \sim p_\theta(\boldsymbol{A}_{re}|\boldsymbol{A})}[-\log q(\boldsymbol{Y}|\boldsymbol{A}_{re})] + \beta \mathrm{KL}[p(\boldsymbol{m}|\boldsymbol{A}),r(\boldsymbol{m})] \qquad (7)$$

where $r(\boldsymbol{m})$ is the prior distribution of $\boldsymbol{m}$ approximating $p(\boldsymbol{m})$; $\boldsymbol{\mu_m}$ and $\boldsymbol{\Sigma}$ are the parameters approximating the multivariate Gaussian distribution of $\boldsymbol{m}$. $\mathcal{L}_{AIB}$ is the optimization target of D-CoRP and $\mathcal{L}_{AIB-E}$ is the optimization target of D-CoRP: Efficient, a lighter version of D-CoRP. The proof can be found in the Supplementary. Since the left-sides of Eq. (6) and (7) are already done by the model comparing the prediction with ground truth, we only optimize an additional $\mathrm{KL}[\mathcal{N}(\boldsymbol{\mathcal{P}};\boldsymbol{\Sigma}),\mathcal{N}(\boldsymbol{\mu_m};\boldsymbol{\Sigma})]$ or $\mathrm{KL}[p(\boldsymbol{m}|\boldsymbol{A}),r(\boldsymbol{m})]$. For D-CoRP: Efficient, the plugin adds an additional $\mathcal{O}(NF^2)$ in computational complexity to the plugged model; For D-CoRP, the additional computational complexity is $\mathcal{O}(N^6)$

	Baseline	Plugin	FCP		NKI-Rockland		ADHD-Normal	
			MAE	RMSE	MAE	RMSE	MAE	RMSE
Basic GNN Module	GCN [12]	None	15.27±1.47	21.20±6.76	85.98±45.23	209.49±127.55	10.57±3.56	18.23±5.51
		D-CoRP (Single)	8.94±0.15	12.13±0.28	16.14±0.36	20.32±1.06	2.77±0.23	3.45±0.22
		D-CoRP (Full)	9.36±0.59	12.53±0.61	16.76±1.22	21.28±1.22	2.96±0.17	3.87±0.24
	GAT [23]	None	20.44±7.03	22.08±6.87	23.36±12.65	28.78±12.44	3.17±0.18	3.59±0.24
		D-CoRP (Single)	9.31±0.16	13.03±0.26	16.06±1.50	20.18±1.51	2.83±0.19	3.47±0.23
		D-CoRP (Full)	9.80±0.74	13.37±1.08	18.00±1.45	22.26±2.70	3.28±0.27	4.20±0.30
	GIN [29]	None	10.85±0.38	14.76±1.71	17.43±4.12	20.28±3.68	5.92±1.47	8.90±3.38
		D-CoRP (Single)	9.11±0.18	12.24±0.38	15.62±1.43	19.60±1.66	2.73±0.13	3.42±0.17
		D-CoRP (Full)	9.29±0.29	12.50±0.24	**15.31±0.73**	**19.06±0.97**	3.18±0.41	3.89±0.46
	GraphSAGE [10]	None	15.72±2.51	30.24±11.99	44.55±14.64	104.86±60.86	4.42±0.26	5.44±0.32
		D-CoRP (Single)	9.28±0.39	12.33±0.47	15.19±0.91	19.71±1.42	**2.65±0.17**	**3.21±0.18**
		D-CoRP (Full)	9.35±0.50	12.70±0.50	17.06±1.77	20.71±2.01	2.87±0.25	3.57±0.30
Basic Graph Pooling	DiffPool [30]	None	12.10±0.79	18.01±4.55	26.96±9.63	39.04±19.44	4.06±0.21	6.56±0.86
		D-CoRP (Single)	**8.55±0.17**	**11.83±0.60**	16.32±2.44	20.39±2.32	2.99±0.20	3.75±0.18
		D-CoRP (Full)	8.84±0.70	11.96±0.73	24.62±10.24	28.56±10.85	3.08±0.26	3.89±0.36
	TopK [9]	None	11.05±0.69	13.23±1.11	27.49±9.42	37.61±10.82	6.14±3.62	6.70±3.42
		D-CoRP (Single)	12.22±6.40	15.21±5.72	19.88±8.72	24.10±9.11	2.73±0.20	3.78±0.10
		D-CoRP (Full)	13.98±4.46	16.34±4.48	30.13±8.31	35.16±9.36	2.84±0.20	3.58±0.28

Table 1: **Performance of D-CoRP on Baselines**. Underlines indicate local optimal comparing None (D-CoRP), D-CoRP (Single) (added to a single GNN block), and D-CoRP (Full) (added to all GNN blocks). Underlines indicate optimal results of corresponding baselines and bold indicates global optimal results.

3 Experiments and Results

3.1 Task and Datasets

We focus on brain age prediction as the primary task. Brain age is shown as associated with brain network features [18,2], serving as an ideal benchmark for model evaluation. The resting-state fMRI data from three datasets processed

with Craddock 200 atlas [7] were used, in which the adjacency matrix consists of Pearson correlation between time series across brain regions and region feature comes from adjacency matrix (each row in the adjacency matrix stands for the feature of corresponding brain region): 1) 1000 Functional Connectomes Project (FCP) [3] (1,003 healthy subjects aged 18-85 yrs (mean 28±13 yrs, 569 females)). 2) Enhanced Nathan Kline Institute - Rockland Sample (NKI-Rockland) (393 subjects, mean 35±20 yrs, 164 females) [16,22], known for wide age range (4-85 yrs). 3) Normal subjects (no diseases) from Attention-Deficit Hyperactivity Disorder-200 (ADHD-Normal) (330 subjects, mean 12±3 yrs, 166 females) [6], focusing on the younger group. All datasets are available in the UCLA multi-modal connectivity database [4].
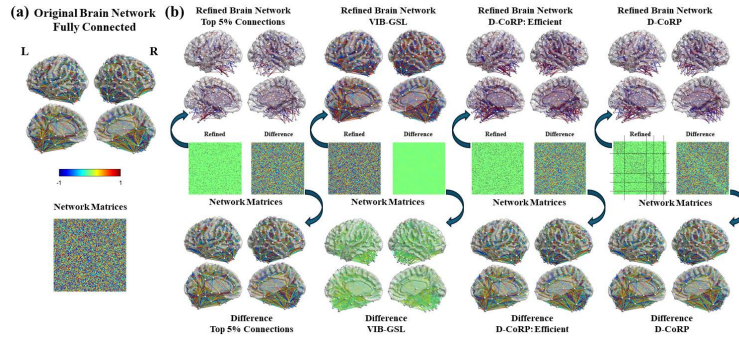

Fig. 2: **Visualization of Refined Brain Network.** One example of the brain networks refined by D-CoRP and other denoising methods. This figure aims to give a visualization of the important connections learned by D-CoRP compared with other methods. (a): original brain network; (b): refined brain networks and their differences from the original network.

3.2   Experiment Settings and Evaluation Metrics

We use two GNN blocks and a multi-layer perceptron (MLP) as the backbone (see Supplementary for detailed structures). We used pooling methods (TopK and DiffPool) with a pooling ratio of 0.25.

In section 3.3, to test the adaptability of D-CoRP, we plug it into 1) the initial part of the model (D-CoRP (Single)), which refines the original brain network; 2) every GNN block of the model (D-CoRP (Full)), which refines the original and processed brain network. The $\boldsymbol{\Sigma}$ is set as $\boldsymbol{I}_{N^2 \times N^2}$, $\boldsymbol{\mu_m}$ is set as $\frac{\mathbf{1}_{N^2}}{20}$, where $\mathbf{1}_{N^2}$ is $N \times N$ matrix with value 1 in each place; $\mathcal{T} = 0.01$, and $\beta = 1$. In section 3.4, for D-CoRP: Efficient, $r(\boldsymbol{m})$ is set as a fixed probability, 0.05.

All training is performed on an Nvidia A-100 with pytorch 2.1 and pytorch-geometric package [8]. We choose Adam as the optimizer with a cosine annealing

learning rate $\frac{1}{10}\cos(\frac{epoch_c}{epoch_T})$, where $epoch_c$ stands for the current epoch number counting from 0 and $epoch_T$ stands for the training epoch set as 150. MSE is used as the loss function. 5-fold cross-validation is applied to all datasets.

Two performance metrics are adopted for evaluation: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), i.e., MAE $= \frac{\sum_{n=1}^{N}|y_i-\hat{y_i}|}{N}$, RMSE $= \sqrt{\frac{\sum_{n=1}^{N}(y_i-\hat{y_i})^2}{N}}$, where y_i is the ground truth and $\hat{y_i}$ is the predicted value.

3.3 Performance of D-CoRP on Baselines

To test the applicability of D-CoRP, we tested the most widely used GNN baselines: GCN [12], GAT [23], GIN [29], and GraphSAGE [10]. Due to the importance of graph pooling methods in brain network research[24,11], we also tested TopK pooling and DiffPool[9,30], known for their capability to distil subgraphs from brain networks.

The experimental results (Table 1) show that D-CoRP significantly outperforms baseline models in all comparisons, supporting the effectiveness of D-CoRP. Of note, with D-CoRP integrated, the backbones exhibit reduced standard deviations in the assessment metrics, suggesting improved stability. This improvement is particularly relevant in brain network research due to data heterogeneity arising from demographics and acquisition. Our results also reveal that adding D-CoRP to all the GNN blocks, i.e., D-CoRP (Full), however, does not perform better in most cases, indicating over-compressed information.

Method	Baseline	FCP			NKI-Rockland			ADHD-Normal		
		MAE	RMSE	Time(s)	MAE	RMSE	Time(s)	MAE	RMSE	Time(s)
	GCN	18.62±4.27	30.22±7.36	170.76	55.35±10.44	82.12±24.45	47.87	11.42±5.86	22.21±6.52	67.72
N/A-5%	GAT	21.30±2.88	46.66±2.98	168.96	44.43±7.46	62.27±8.69	53.32	16.32±2.42	20.78±3.11	69.91
	GIN	21.80±2.32	25.99±2.57	171.49	30.69±3.46	38.33±4.21	55.62	9.45±2.95	15.74±4.01	70.10
	GCN	27.24±0.15	30.06±0.20	602.43	33.96±1.72	39.39±1.46	202.32	11.56±0.22	12.01±0.23	331.29
SIB [31]	GAT	27.41±0.22	30.21±0.41	610.77	35.62±0.92	41.18±1.23	209.12	11.80±0.43	12.25±0.98	347.55
	GIN	27.28±1.25	29.81±1.31	609.72	35.40±1.66	40.80±2.53	204.86	11.23±1.45	11.67±1.64	354.88
	GCN	**8.66±0.40**	13.90±0.58	570.16	16.44±1.88	21.02±1.63	110.91	4.66±1.68	5.73±1.54	180.77
VIB-GSL [19]	GAT	11.10±0.94	17.10±0.85	578.66	16.69±1.43	22.01±1.43	114.86	4.55±0.76	5.55±0.75	182.63
	GIN	12.54±0.46	17.78±0.61	570.70	30.58±1.71	36.33±2.02	114.10	8.92±0.80	9.66±0.79	180.75
	GCN	11.50±0.96	16.66±0.98	**181.66**	34.23±1.33	41.07±3.17	**54.43**	5.57±1.36	10.68±2.01	**78.20**
D-CoRP: Efficient	GAT	12.96±2.07	20.41±4.65	**182.53**	36.71±1.93	41.53±2.42	**60.55**	**2.53±0.41**	**2.96±0.44**	**82.22**
	GIN	11.76±1.08	16.22±1.16	**188.34**	16.59±0.77	20.00±1.00	**56.33**	9.66±2.01	15.93±3.08	**83.12**
	GCN	8.94±0.15	**12.13±0.28**	579.44	16.14±0.36	20.32±1.06	132.37	2.77±0.23	3.45±0.22	228.67
D-CoRP	GAT	9.31±0.16	13.03±0.26	580.87	16.06±1.50	20.18±1.51	133.24	2.83±0.19	3.47±0.23	231.18
	GIN	9.11±0.18	12.24±0.38	586.53	**15.62±1.43**	**19.60±1.66**	133.10	2.73±0.13	3.42±0.17	222.51

Table 2: **Comparison of D-CoRP and Other Denoising Methods.** The bold font indicates the global optimal results for each dataset on MAE and RMSE metrics. The time metric in the table shows the time to run the model on 5-fold cross-validation, and the bold font shows optimal results excluding N/A-5%.

3.4   Comparison with Other Denoising Methods

We compare D-CoRP and D-CoRP: Efficient with conventional denoising method, selecting top 5% connections according to correlation values. We also compare two IB-based graph structure learning methods, i.e., SIB [31], and VIB-GSL [19]. All comparisons involve three different backbones (i.e., GCN, GAT and GIN) as the graph processing module.

From Table 2, we observe that D-CoRP outperforms other models in most comparisons. In all cases, D-CoRP shows better stability over the comparison models, indicated by lower variance. These results support the effectiveness of D-CoRP for brain network research. Additionally, our results imply that D-CoRP: Efficient may be more suitable for smaller datasets (e.g., ADHD-Normal).
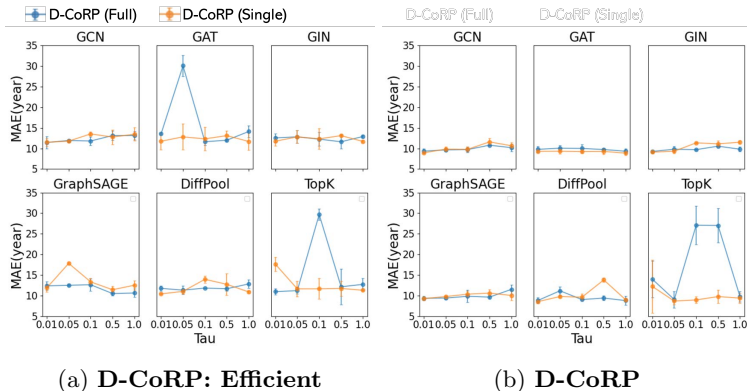

(a) **D-CoRP: Efficient**   (b) **D-CoRP**

Fig. 3: **Results of Variant Temperature.**

Further, we visualize an example network derived from the ADHD-Normal dataset in Fig. (2). The results show that our methods effectively remove noisy connections compared to other methods. Interestingly, from the refined network from D-CoRP, we observe the modularity preserved in the functional network, one of the most important brain network properties, suggesting the D-CoRP's capability to preserve informative edges alongside denoising.

3.5   Hyperparameter Tuning on D-CoRP

The temperature $\mathcal{T}$ controls the slope of sampling function $\mathcal{F}(\mathcal{P}, \boldsymbol{\Pi} : \mathcal{T})$, influencing how effectively D-CoRP refines brain network connectivities. Six backbone models in Table. 1 are utilized to test the effectiveness of our D-CoRP. 10-fold cross-validation is performed on the FCP dataset. MAE is used to evaluate the final performance. Fig. 3 shows $\mathcal{T}$ value of 0.01 for D-CoRP may lead to superior and stable model performance in more than half of cases and setting $\mathcal{T}$ as 1 may be generally suitable for all cases. We recommend setting $\mathcal{T}$ as 0.01

if only binary masking is needed while setting $\mathcal{T}$ as 1 if D-CoRP is also utilized to adjust the importance of connections in addition to the denoised part.

4   Conclusion

In this study, we propose D-CoRP, a tool for differentiable connectivity refinement in functional brain networks based on information bottleneck theory. Specifically, we design a differentiable sampling function and multivariate probability modeling for detecting valuable connections in the brain network. Our experiments show that D-CoRP achieves state-of-the-art performance in connectivity refinement. As a plugin, D-CoRP can be integrated into existing GNN-based frameworks and provides an efficient mode for flexible adoption. Future work will develop multivariate optimization with adaptive covariance tailored to functional connectivity.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alemi, A.A., Fischer, I., Dillon, J.V., Murphy, K.: Deep variational information bottleneck. arXiv preprint arXiv:1612.00410 (2016)
2. Bashyam, V.M., Erus, G., Doshi, J., Habes, M., Nasrallah, I.M., Truelove-Hill, M., Srinivasan, D., Mamourian, L., Pomponio, R., Fan, Y., et al.: Mri signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide. Brain **143**(7), 2312–2324 (2020)
3. Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., et al.: Toward discovery science of human brain function. Proceedings of the national academy of sciences **107**(10), 4734–4739 (2010)
4. Brown, J.A., Rudie, J.D., Bandrowski, A., Van Horn, J.D., Bookheimer, S.Y.: The ucla multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. Frontiers in neuroinformatics **6**, 28 (2012)
5. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews neuroscience **10**(3), 186–198 (2009)
6. consortium, A..: The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Frontiers in systems neuroscience **6**, 62 (2012)
7. Craddock, R.C., James, G.A., Holtzheimer III, P.E., Hu, X.P., Mayberg, H.S.: A whole brain fmri atlas generated via spatially constrained spectral clustering. Human brain mapping **33**(8), 1914–1928 (2012)
8. Fey, M., Lenssen, J.E.: Fast graph representation learning with pytorch geometric. arXiv preprint arXiv:1903.02428 (2019)
9. Gao, H., Ji, S.: Graph u-nets. In: international conference on machine learning. pp. 2083–2092. PMLR (2019)

10. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. Advances in neural information processing systems **30** (2017)
11. Hilgetag, C.C., Goulas, A.: 'hierarchy'in the organization of brain networks. Philosophical Transactions of the Royal Society B **375**(1796), 20190319 (2020)
12. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
13. Levakov, G., Rosenthal, G., Shelef, I., Raviv, T.R., Avidan, G.: From a deep learning model back to the brain—identifying regional predictors and their relation to aging. Human brain mapping **41**(12), 3235–3252 (2020)
14. Li, C., Wei, Y., Chen, X., Schönlieb, C.B.: Brainnetgan: Data augmentation of brain connectivity using generative adversarial network for dementia classification. In: Deep Generative Models, and Data Augmentation, Labelling, and Imperfections: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 1. pp. 103–111. Springer (2021)
15. Li, X., Zhou, Y., Dvornek, N., Zhang, M., Gao, S., Zhuang, J., Scheinost, D., Staib, L.H., Ventola, P., Duncan, J.S.: Braingnn: Interpretable brain graph neural network for fmri analysis. Medical Image Analysis **74**, 102233 (2021)
16. Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L., Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., et al.: The nki-rockland sample: a model for accelerating the pace of discovery science in psychiatry. Frontiers in neuroscience **6**, 152 (2012)
17. Park, H.J., Friston, K.: Structural and functional brain networks: from connections to cognition. Science **342**(6158), 1238411 (2013)
18. Stanley, M.L., Simpson, S.L., Dagenbach, D., Lyday, R.G., Burdette, J.H., Laurienti, P.J.: Changes in brain network efficiency and working memory performance in aging. PLoS One **10**(4), e0123950 (2015)
19. Sun, Q., Li, J., Peng, H., Wu, J., Fu, X., Ji, C., Philip, S.Y.: Graph structure learning with variational information bottleneck. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 4165–4174 (2022)
20. Tishby, N., Pereira, F.C., Bialek, W.: The information bottleneck method. arXiv preprint physics/0004057 (2000)
21. Tishby, N., Zaslavsky, N.: Deep learning and the information bottleneck principle. In: 2015 ieee information theory workshop (itw). pp. 1–5. IEEE (2015)
22. Tobe, R.H., MacKay-Brandt, A., Lim, R., Kramer, M., Breland, M.M., Tu, L., Tian, Y., Trautman, K.D., Hu, C., Sangoi, R., et al.: A longitudinal resource for studying connectome development and its psychiatric associations during childhood. Scientific Data **9**(1), 300 (2022)
23. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
24. Vidaurre, D., Smith, S.M., Woolrich, M.W.: Brain network dynamics are hierarchically organized in time. Proceedings of the National Academy of Sciences **114**(48), 12827–12832 (2017)
25. Wang, P.Y., Sapra, S., George, V.K., Silva, G.A.: Generalizable machine learning in neuroscience using graph neural networks. Frontiers in artificial intelligence **4**, 618372 (2021)
26. Wei, Y., Chen, X., Zhu, L., Zhang, L., Schönlieb, C.B., Price, S., Li, C.: Multimodal learning for predicting the genotype of glioma. IEEE Transactions on Medical Imaging (2023)

27. Wei, Y., Li, C., Price, S.J.: Quantifying structural connectivity in brain tumor patients. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 519–529. Springer (2021)
28. Williams, N., Arnulfo, G., Wang, S.H., Nobili, L., Palva, S., Palva, J.M.: Comparison of methods to identify modules in noisy or incomplete brain networks. Brain connectivity **9**(2), 128–143 (2019)
29. Wu, T., Ren, H., Li, P., Leskovec, J.: Graph information bottleneck. Advances in Neural Information Processing Systems **33**, 20437–20448 (2020)
30. Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., Leskovec, J.: Hierarchical graph representation learning with differentiable pooling. Advances in neural information processing systems **31** (2018)
31. Yu, J., Xu, T., Rong, Y., Bian, Y., Huang, J., He, R.: Graph information bottleneck for subgraph recognition. arXiv preprint arXiv:2010.05563 (2020)
32. Zhang, P., Wen, G., Cao, P., Yang, J., Zhang, J., Zhang, X., Zhu, X., Zaiane, O.R., Wang, F.: Brainusl: U nsupervised graph s tructure l earning for functional brain network analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 205–214. Springer (2023)
33. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., Sun, M.: Graph neural networks: A review of methods and applications. AI open **1**, 57–81 (2020)
34. Zhu, Y., Xu, W., Zhang, J., Liu, Q., Wu, S., Wang, L.: Deep graph structure learning for robust representations: A survey. arXiv preprint arXiv:2103.03036 **14** (2021)